PERSPECTIX

NEW YORK UNIVERSITY

Analysing NYC Taxi Trends To Understand Humans Of New York

Abhishek Pillai Kunal Relia Sanjitha Udipi

Background and Motivation

"Data doesn't speak for itself, we have to make it speak."

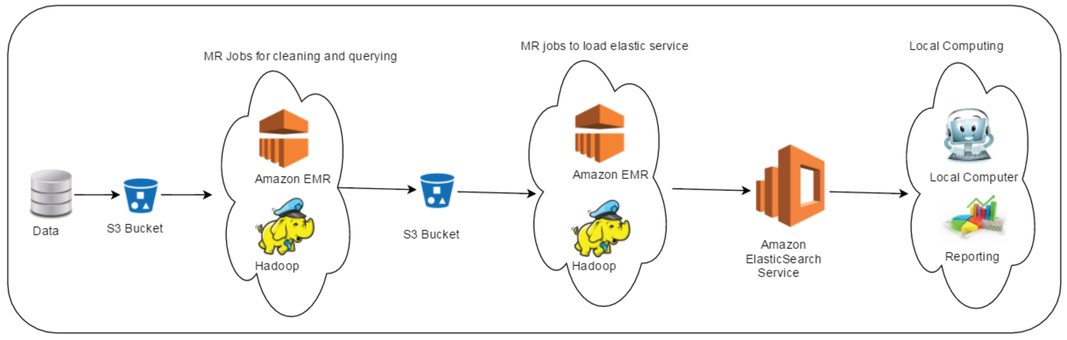
Based on this famous quote by a well-known Data Scientist, we decided to explore the NYC taxi trends to understand the New Yorkers better. The NYC taxis make millions of trips each year and a lot of information is hidden in the data gathered from it. Hence, we explored the pattern in which New Yorkers take taxi rides and pay for the rides. The hidden information will enable us to get them know better.

Data and Method

The taxi ***Trips*** and ***Fares*** data used for analysis was taken from the project done by Dan Work.

The Trips set contained the pickup and dropoff locations and time stamps along with the trip duration. The Fares table contained each trip's monetary parameters like various amount and payment method. The ***Weather*** data provided by NOAA would help in interpreting the change in the usage of taxi change in temperatures.

Model and Overview



Results

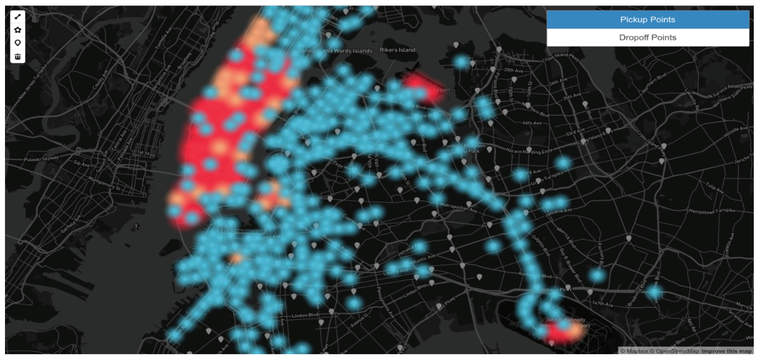


Fig A: The Heat map depicts the ***most popular pickup locations***. As expected, ***Manhattan and the two airports*** have the highest pickups in NYC.

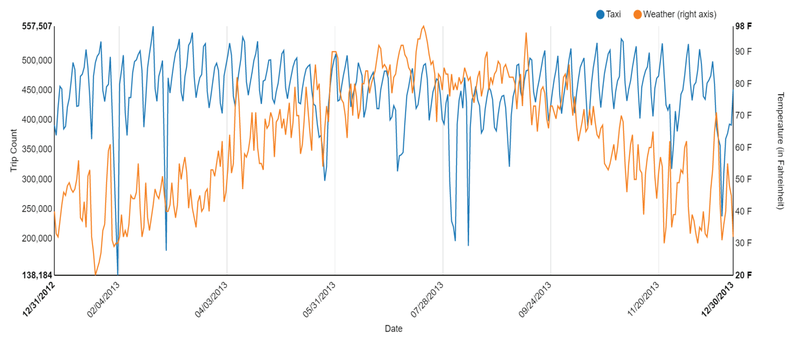
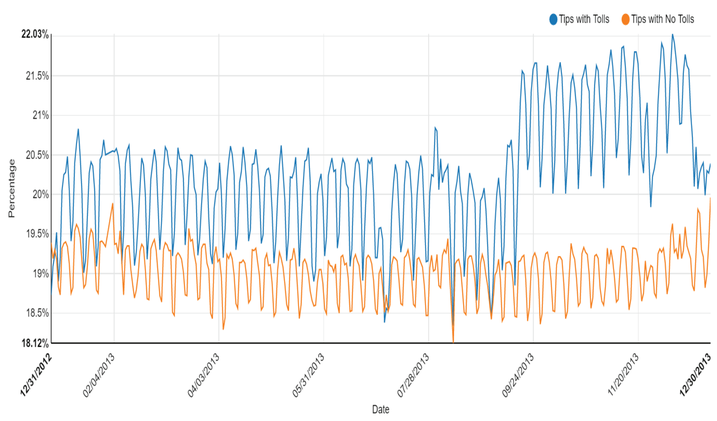


Fig B: The multi-line chart shows the relation between the ***temperature and the number of trips.*** In most cases, there is an increase in trips with a fall in temperature.



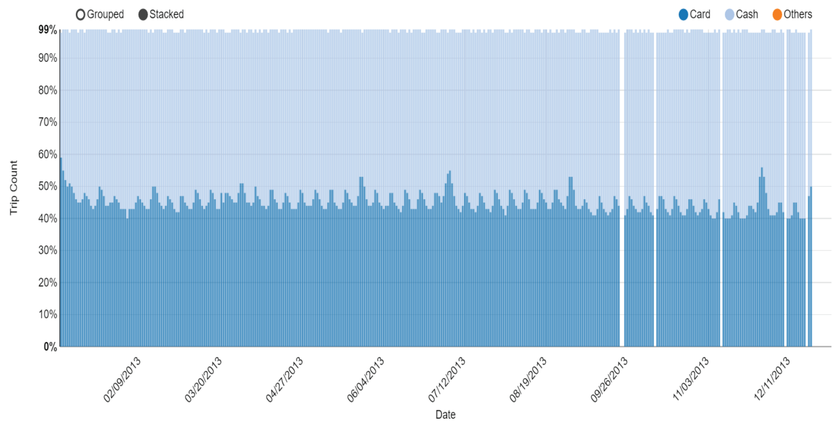


Fig C: The multi-line chart shows the relation percentage of tip paid when there were trips involving toll tax and trips not involving toll tax. It turns out that as the average distance of a trip having a toll tax is large, the percentage of tip paid is higher. This means New Yorkes ***pay more tips on longer trips***. The sinosudal nature of the line chart is due to an interesting characteristic of New Yorkers paying ***higher tips during the weekdays*** and relatively ***lower tips on the weekends.*** (Only trips paid with card have been analyzed as it is found that many tips paid in cash might not have been recorded.

Fig D: Shows the likeliness of a commuter using a ***card to pay the fare over cash***.

Conclusion

The taxi use over the years reveal a lot about the people using it. Since the release of NYC taxi data, it has been a favorite of all big data players. After a lot of research done and papers written, a lot still remains unexplored. There have been similar talks as ours on taxi usage over a period. But we feel that the fares have been a neglected dataset.

In future, we can build upon this idea of Exploring NYC taxi data to understand Humans of New York to study the behavior of New Yorkers. Smallest information like the tips paid or the different clustering of pickup points over weekdays and weekends speak volumes about people staying here. More analysis can bring in more interesting facts out of the bag.

Acknowledgements

We thank: ***Prof. Juliana Freire*** of New York University for guiding us and giving insights through the project.

Dr.Erin Carson and ***Dr.Nick Knight*** of New York University for their overall support and suggestions.

Prof. Dan Work of University of Illinois - Urbana-Champaign for Taxi Data and ***NOAA*** for Weather Data

Amazon for Education Coursework grant to use their cloud services